Application for a PhD Grant Programme

at the Laboratory of Excellence UCN@SOPHIA

June 2013

Title of the PhD Project: Random matrix analysis for distributed algorithms and complex networks

Involved Research Units: Inria and EURECOM

PhD Advisors: Konstantin Avrachenkov (Maestro team), Laura Cottatellucci (Mobile Communications)

Abstract

The feverish growth of the Internet, social networks, the quest for a deep understanding of logical, biological and physical networks for nanotechnologies is fuelling intense researches to investigate fundamental interconnection structures and dynamics of physical and logical complex networks and gave birth to a new discipline called "Network Science". Network Science aims to develop a new theory, mathematical principles and algorithms to understand the intrinsic communication patterns, the growth and decay dynamics, as well as information propagation, and distributed algorithms in such complex networks. A complex network can be effectively modeled by random matrices. The properties of the random ensemble characterize a complex network as a whole, beyond the peculiarities of single realizations. We intend to gain insights on complex networks via Random Matrix Theory (RMT) and use them to develop synergies with adjacent research fields. More specifically, we aim to apply the general results gained via RMT to (a) the design and analysis of distributed randomized algorithms for very large networks, (b) to the study of information search and classification, and (c) task distribution and resource allocation among nodes and clusters of nodes in the Cloud and in particular in C-RAN.

Context and the State of Art

"Network Science" is a new emerging scientific discipline that examines the fundamental interconnection structures and dynamics of physical and logical complex networks. Examples of physical networks studied in network science include the Internet, biological networks, and transportation networks. Examples of logical networks are BitTorrent overlay networks, P2P networks like Skype, online social networks like Facebook, Servers' Network in the Cloud as well as web networks which interconnect web pages around the world. There is an urgent need to develop new theory, mathematical principles, and algorithms to understand the intrinsic communication patterns, the growth and decay dynamics, as well as information propagation in such networks. Furthermore, these networks are typically gigantic in size, e.g., Facebook is experiencing an exponential growth and consists of around 1.11 billion of nodes (as of March 2013) and billions of connections. Twitter, another popular online social network, has over 500 million registered users as of 2012. It is fundamental to analyze these large dynamic networks and understand their topological and dynamical properties (e.g. connectivity, clustering, assortativity, diffusion processes on networks, etc.). There are many potential applications in this new discipline, for example, (a) task distribution and resource allocation among nodes and clusters of nodes in the Cloud and in particular in C-RAN (Cloud Radio Access Network), (b) information search and classification, (c) understanding of the epidemic spreading of computer virus and data pollution in large networks so to provide appropriate countermeasures, (d) analysis of road networks and traffic bottlenecks.

Complex networks can be effectively modeled by random graphs. The first fundamental random graph model for complex networks, the *Erdos-Renyi model*, was introduced by Erdos and Renyi in their seminal work [ER59] and it can be described by independently connected each other some given probability. Nowadays, intensive efforts are devoted to the development of more structured models. Recently, the *configuration model* has been proposed (see e.g. [NWS01],[N03]). In this case, the network is characterized by a given distribution of the node degrees. This model can be mapped onto sparse matrices with a given description of sparsity. The *generative model* based on Kronecker matrix multiplication recently proposed by Leskovec et al. [LCKF05] is a model that captures many properties of real-world networks.

The possibility to describe the topological properties of complex networks via spectral properties of random graphs has been widely explored in different scientific communities via empirical simulations (see e.g. [FDBTV01], [BJ07], [MT09], [SKHB05]). An exhaustive overview of the literature in this stream exceeds the scope of this project proposal. A deep relation between random graphs and random matrices is apparent. Let take for example a particular graph instance. It can be described by its adjacency matrix. An element (i,j) of the adjacency matrix is equal to one if there is a link from node i to node j and zero otherwise. Similarly, a random graph can be described by a random matrix. Nevertheless, the interactions between the two scientific communities working on these two topics have been very limited.

The adjacency matrix spectrum of the Erdos-Renyi random graph corresponds to one of the most deeply studied classes of random matrices in RMT: the Wigner matrices. Therefore, initial analytical results on random matrices for complex networks are the analysis of the Erdos-Renyi model by making use on classical existing results on the eigenvalue distribution of Wigner matrices and its greatest eigenvalue (see e.g. Chapter 6 in [M11] for a review). Very recent and powerful results on the local eigenvalue statistics of generalized Wigner matrices provide a finer description of their spectral properties via the eigenvalue gap distribution [TV11]. Oriented Erdos-Renyi random graphs can be modeled by random matrices with independent and identically distributed entries, a class of matrices also deeply studied, whose properties and universality, for very long time only conjectured, has been recently proven in [TVK10]. For the characterization of random walks in graphs, a prominent role is played by the Markov and Laplacian matrices associated to the adjacency matrix of the graph. The spectral analysis of

Markov and Laplacian matrices of a sparse oriented Erdos-Renyi random graph has been recently proposed in [BCC12a,BCC12b]. In this case, a key role is played by free probability for quaternions [JN06]. These promising, however largerly incomplete, results are fostering interesting studies on complex networks based on RMT. This research field is still in its infancy.

Many existing algorithms for network analysis are mainly for in-memory processing and limited to millions of nodes at best. Therefore, it is imperative to develop methods with quasi-linear, linear, and even sub-linear complexity for the efficient analysis of very large networks (hundreds of millions or even several billion nodes). One very promising research direction is the development of randomized algorithms or Monte Carlo algorithms, which trade accuracy for execution time [MR95]. In fact, Monte Carlo type algorithms are particularly well-suited for parallel distributed implementations. We have already some initial results in this research direction with respect to quick detection of the most central nodes in complex networks [ALNO07,ALNSS11]. An overview of various centrality measures can be found in [BE06] (see also references therein). In [ADNPS08] we propose a PageRank based method for clustering large hypertext document collections which has quasi-linear computational complexity. Spectral properties of the underlying graph play a key role in the performance analysis and design of Monte Carlo type algorithms.

Challenges

Network science has already a significant impact on various scientific disciplines such as biology, physics, communications, sociology. However, up to the present, the most of the results in network science are obtained by empirical measurements and simulations even limited in scale. Clearly, a fundamental basis of network science is still missing. This proposal aims to contribute to the creation of a solid theoretical background in network science.

The traditional algorithms for graph analysis become just infeasible for large complex networks. Even algorithms with quadratic complexity cannot be executed in reasonable time for networks with the order of 100,000 nodes. The eigenvalue spectrum of a complex network provides fundamental information about its structural properties. However, its numerical computation has a prohibitive complexity, both in time and memory, even with few thousands of nodes, and constitutes a formidable challenge in the analysis of complex networks. Therefore, it is imperative to develop methods with quasi-linear, linear and even sub-linear complexity for an efficient analysis of very large networks.

Additionally, a numerical analysis does not provide very insightful results about the dependence of the network properties on macroscopic network parameters and has a limited usefulness in the system tuning and shaping.

Also the applications related to the proposed research topic present exciting challenges due to scalability issues. Information retrieval applications create a huge amount of interlinked data. The air traffic and car usage is still rapidly growing worldwide. Clustering of front end devices in C-RAN infrastructure plays a key role in network scalability. New strains of viruses and fishing constantly challenge computer systems and threaten users.

Expected Results

Random matrix theory (RMT) is a branch of mathematics that investigates the spectral properties of random matrices. It has been successfully applied to many disciplines where complex interactions among a large number of entities play a key role. The spectra of many classes of random matrices converge to asymptotic deterministic limits as the size of the matrix grows large and the spectra can be analytically expressed in terms of few parameters. As a consequence, the properties of complex systems related to these random matrix spectra can also be expressed in terms of few macroscopic system parameters. Very representative and fruitful applications of RMT have been the analysis and the design of multiuser, multi-antenna wireless networks. In such a context, Gramian random matrices with independent and/or correlated entries play a fundamental role and have been object of thorough studies also by the project proposing team (see e.g. [TV04], [CM05], [CM07], [CMD10a], [CMD10b]). Interestingly, fundamental properties and limits of wireless networks could be expressed in terms of macroscopic system parameters and fundamental insights have been gained by using tools of RMT.

It seems extremely promising to investigate complex networks by a similar approach. Indeed, complex networks can be effectively described by statistical ensembles of matrices. Thus, when we consider the extreme difficulties in measuring, analyzing characteristics of complex networks, and design efficient distributed algorithms over complex networks, the benefit of having such an analytical framework becomes apparent.

The aim of our research is to extend the benefits that RMT had in telecommunication systems to complex networks and derive analytical results on the spectra of large matrices modeling complex networks. This will allow us to establish fundamental relations between macroscopic parameters and topological properties of the networks. The current intensive efforts devoted to the development of realistic models imply the introduction of new statistical classes of random matrices. Each class requires an independent analysis via RMT. In this PhD project, the candidate will develop or extend methods of RMT to the analysis of random matrices modeling complex networks. In such a way, she/he will contribute to the construction of theoretical foundations of network science. We plan to study and analytically characterize principal eigenvalues and to investigate non-principal eigenvalues in complex network evolution. For example, the spectral gap of the network Laplacian matrix determines the rate of convergence of a random walk to its stationary distribution. Consequently, this will contribute to the performance analysis of randomized algorithms. In this way, we already contribute to two important areas of application for complex networks: design and analysis of scalable and distributed algorithms for very large networks and information propagation in complex networks.

As applications, we also plan to develop efficient methods for quick detection of important nodes according to various centrality measures such as in and out degrees, betweenness centrality, closeness centrality. Our research effort in using RMT will provide a valuable tool to characterize the rate of algorithms convergence for various typical graphs. Although there is a number of studies on graph clustering or community detection, only very few algorithms can work with very large graphs. Even less research has been done for clustering large dynamic graphs. We intend to propose efficient random walk based methods for clustering, centrality measures and connectivity assessment of dynamic graphs. We also propose to use the random walk approach to develop new distributed search algorithms which perform well on most popular types of complex networks.

References

[ALNO07] K. Avrachenkov, N. Litvak, D. Nemirovsky and N. Osipova, Monte Carlo methods in PageRank computation: When one iteration is sufficient, SIAM Journal on Numerical Analysis, v.45, no.2, pp.890-904, 2007.

[ALNSS11] K. Avrachenkov, N. Litvak, D. Nemirovsky, E. Smirnova and M. Sokol, Quick Detection of Top-k Personalized PageRank Lists, LNCS v.6732, pp.50-61, 2011.

[ADNPS08] K. Avrachenkov, V. Dobrynin, D. Nemirovsky, S.K. Pham and E. Smirnova, PageRank Based Clustering of Hypertext Document Collections . Proceedings of ACM SIGIR 2008.

[BCC12a] C. Bordenave, P. Caputo and D. Chafaı, “Circular Law Theorem for Random Markov Matrices”, Probability Theory and Related Fields, vol 152, pp. 751-779, 2012.

[BCC12b] C. Bordenave, P. Caputo, and D. Chafai. "Spectrum of Markov generators on sparse random graphs." arXiv preprint arXiv:1202.0644 (2012).

[BJ07] Jayendra N. Bandyopadhyay, and Sarika Jalan, Universality in complex networks: Random matrix analysis, Phys. Rev. E 76, 026109, 2007

[BE06]S.P. Borgatti, M.G. Everett, ``A Graph-theoretic perspective on centrality’’, Social Networks, Volume 28, Issue 4, October 2006, Pages 466-484.

[CMD10a] Laura Cottatellucci, Ralf R. Müller, and Merouane Debbah. “Asynchronous CDMA systems with random spreading–Part I: Fundamental limits”. IEEE Transactions on Information Theory, vol. 56, no. 4, pp. 1477 - 1497 Apr 2010.

[CMD10b] Laura Cottatellucci, Ralf R. Müller, and Merouane Debbah. Asynchronous CDMA systems with random spreading–Part II: Design criteria. IEEE Transactions on Information Theory, vol. 56, no. 4, pp. 1498 – 1520 Apr 2010.

[CM07] Laura Cottatellucci and Ralf R. Müller. “CDMA Systems with correlated spatial diversity: A generalized resource pooling result”. IEEE Transactions on Information Theory, vol. 53, no. 3, pp. 1116 - 1136, Mar 2007.

[CM05] Laura Cottatellucci and Ralf R. Müller. “A systematic approach to multistage detectors in multipath fading channels”. IEEE Transactions on Information Theory, vol. 51, no. 9, pp. 3146–3158, Sep 2005.

[CD11] Couillet, Romain, and Merouane Debbah. Random matrix methods for wireless communications. Cambridge University Press, 2011.

[ER59] P. Erdös, A. Rényi, On random graphs, I, Publicationes Mathematicae (Debrecen), Vol. 6 (1959), pp. 290-297

[FDBV01] Farkas, Derényi, Barabási, and Vicsek, Spectra of "real-world" graphs: Beyond the semicircle law, Physical Review E 64, 026704, 2001.

[GKK01] K.-I. Goh, B. Kahng, and D. Kim. Spectra and Eigenvectors of Scale-Free Networks. Physical Review E, 64:051903, 2001.

[JN06] Jarosz, A., and M. A. Nowak. "Random Hermitian versus random non-Hermitian operators—unexpected links." Journal of Physics A: Mathematical and General 39 (2006): 10107.

[LCKF05] J. Leskovec, D. Chakrabarti, J. Kleinberg, and C. Faloutsos. Realistic, mathematically tractable graph generation and evolution, using Kronecker multiplication. PKDD, 2005.

[M11] Piet van Mieghem, Graph spectra for complex networks, Cambridge University Press, 2011.

[MT09] Marija Mitrović and Bosiljka Tadić, Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities, Physical Review E 80, 026123, 2009.

[NN] Raj Rao Nadakuditi and Mark E.J. Newman, "Spectra of random graphs with arbitrary expected degrees". http://arxiv.org/abs/1208.1275v1, Aug 2012.

[NR08] Taro Nagao and G. J. Rodgers, Spectral density of complex networks with a finite mean degree, Journal of Physics A: Mathematical and Theoretical, 2008 41 265002 doi:10.1088/1751-8113/41/26/265002

[NSW01] Newman, Strogatz and Watts, Random graphs with arbitrary degree distributions and their applications, Physical Review E 64, 026118, 2001

[N03] Newman, The structure and function of complex networks. SIAM Review, 45(2):167–256, 2003

[RN10] G. J. Rodgers and Taro Nagao, Complex networks, To appear in Akemann G, Baik J, Di Francesco P. (eds.), Oxford Handbook of Random Matrix Theory, published by Oxford University Press, Publication Date: 2010.

[SKHB05] M. Sade, T. Kalisky, S. Havlin, and R. Berkovits Localization transition on complex networks via spectral statistics, Physical Review E 72, 066123, 2005.

[TV04] Antonia Tulino and Sergio Verdu, Random Matrix Theory and Wireless Communications, in Foundations and Trends in Communications and Information Theory, Vol. 1, publisher NOW, June 2004.

[TVK10] Tao, Terence, Van Vu, and Manjunath Krishnapur. "Random matrices: Universality of ESDs and the circular law." The Annals of Probability 38, no. 5 (2010): 2023-2065.

[TV11] Tao, Terence, and Van Vu. "Random matrices: Universality of local eigenvalue statistics." Acta mathematica 206, no. 1 (2011): 127-204.